Sankhyā: The Indian Journal of Statistics
2000, Volume 62, Series A, Pt. 1, pp. 23–35

HIDDEN TRUNCATION MODELS

By BARRY C. ARNOLD
and
ROBERT J. BEAVER
University of California, Riverside

SUMMARY. Suppose $(\underline{X}, Y)$ has a $k+1$ dimensional normal distribution. Consider the conditional distribution of $\underline{X}$ given $Y > y_0$, for some fixed value $y_0 \epsilon \mathbf{R}$. Such hidden truncation models provide a flexible family of skewed alternatives to the classical k dimensional normal distribution. Distributional properties of these models are investigated. Non-normal variants of the distribution are also discussed as are multiple hidden truncation models. A specific example involving a skewed bivariate data set (heights and weights of athletes) is analysed in detail.

1. **Introduction**

Skewed multivariate distributions can arise in situations in which the observed variables represent a sample that has been truncated with respect to some hidden covariable. An example which readily comes to mind would involve the distribution of weights of police officers who have been selected only if they meet a minimal height requirement. In general one might expect a bivariate normal distribution of heights and weights. However the distribution of weights of above average height individuals will be a skewed non-normal distribution. The density function of such hidden tructation models, was discussed in some detail in Arnold *et al.* (1993) and in Azzalini (1986). The special case in which observations are retained only if they are above average with respect to the hidden variable was discussed also by Azzalini (1985). After location and scale transformations the Azzalini skew-normal density assumes the form

$$f(x; \lambda) = 2\varphi(x)\Phi(\lambda x), \quad x \,\epsilon\, \mathbf{R} \qquad \ldots (1.1)$$

where φ and Φ denote the standard normal density and distribution function and $\lambda \,\epsilon\, \mathbf{R}$ is a "skewness" parameter. Recently Azzalini and Dalla Valle (1996) have introduced a multivariate extension of the model (1.1). After a linear transformation the basic Azzalini-Dalla Valle k-dimensional distribution takes the form

$$f(\underline{x}; \underline{\lambda}) = 2\left[\prod_{i=1}^{k}\varphi(x_i)\right]\Phi\left(\sum_{i=1}^{k}\lambda_i x_i\right), \quad \underline{x} \,\epsilon\, \mathbf{R}^k \ . \qquad \ldots (1.2)$$

Paper received. May 1999; revised November 1999.
AMS (1991) *subject classification.* 62H05; 62H12.
Key words and phrases. Skewed-normals, truncated normals, generalized skewed distributions.

Here $\underline{\lambda}$ is a k-dimensional skewness parameter vector. The model (1.2) will, as we shall see, be equivalent to the one in which, beginning with a $(k+1)$ dimensional normal random vector $(\underline{X}, Y)$ we retain observations iff their Y values are above average. The distribution (1.2) thus is a natural extension of the univariate distribution (1.1). However, it is more likely that the truncation on the hidden variable Y will be at a level not equal to its mean. It is natural to seek a multivariate extension of the Arnold *et al.* (1993) univariate model which allowed general truncation with respect to the hidden variable. As we shall see, the end result of this exercise is a k-dimensional density which is closely related to (1.2) but has one additional "skewness" parameter. After affine transformations the joint density is of the form

$$f(\underline{x}; \lambda_0, \underline{\lambda}) = \left[\prod_{i=1}^{k} \varphi(x_i)\right] \Phi(\lambda_0 + \sum_{i=1}^{k} \lambda_i x_i) \Big/ \Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}'\underline{\lambda}}}\right), \quad x \in \mathbf{R}. \quad \ldots (1.3)$$

Distributional properties of this family of distributions will be discussed in section 4.

It is possible to consider variants of the hidden truncation model (1.3). One might consider cases in which truncation involves more than one hidden variable attaining an acceptable threshhold value, instead of just one (e.g. sitting and standing height instead of just height for our police officers). Such multiple constraint models are discussed briefly in section 6. A second variation involves the use of a non-normal density ψ_1 and a non-normal distribution Ψ_2 in place of the standard normal φ and Φ in the development leading to (1.3). Some results in this direction, focussing on extensions of Azzalini and Dalla Valle's model (1.2), have been described in Arnold and Beaver (1997a). Section 5 of the present paper deals with non-normal extensions of the model (1.3) with special attention paid to the skewed Cauchy case. In section 7, we reevaluate the Australian athletes data, previously fitted by Azzalini and Dalla Valle using a model based on (1.2). The additional skewness parameter present in (1.3) will be shown to lead to a somewhat improved fit of the data. The suggestion is that the hidden truncation for the Australian athletes involved some other threshhold for the hidden variable than just "above average".

We will begin with a brief review of the one dimensional normal case, discussed in detail in Azzalini (1985, 1986), Cartinhour (1990) and Arnold *et al.* (1993).

2. Univariate Hidden Truncation Normal Models

We begin with (X, Y) having a bivariate normal distribution with mean vector (μ_1, μ_2) and variance-covariance matrix $(\sigma_{ij})_{i=1,j=1}^{2,2}$. We then set $Z = X$ conditional on $Y > a$. Azzalini (1985) considered the case in which $a = \mu_Y$, while Arnold *et al.* (1993) considered the more general case. Integration (for details see Arnold *et al.* (1993)) yields

$$f_Z(z) = \frac{1}{\sigma_1} g\left(\frac{z-\mu_1}{\sigma_1}\right) \quad \ldots (2.1)$$

where

$$g(w) = \frac{\varphi(w)\Phi\left(\frac{\rho w-\alpha}{(1-\rho^2)^{1/2}}\right)}{\Phi(-\alpha)} \ , \qquad \ldots (2.2)$$

ρ denotes the correlation in the original (untruncated) population and $\alpha = \frac{a-\mu_Y}{\sigma_Y}$. The parameters ρ and α control the skewness in the basic density $g(w)$.

For purposes of highlighting the relationship between the density (2.2) and Azzalini's (1985) skew-normal density it is convenient to reparameterize as follows:

$$\lambda_0 = -\alpha/\sqrt{1-\rho^2} \qquad \ldots (2.3)$$

$$\lambda_1 = \rho/\sqrt{1-\rho^2} \ . \qquad \ldots (2.4)$$

Both λ_0 and λ_1 can assume any real value. In terms of these parameters, our basic hidden truncation density is given by

$$f(w;\lambda_0,\lambda_1) = \varphi(w)\Phi(\lambda_0+\lambda_1 w)\Big/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right) \ . \qquad \ldots (2.5)$$

Azzalini (1985) mentions a density essentially equivalent to (2.5) in a Bayesian context. (See his equation (9)). An alternative (equivalent) formulation of the hidden truncation model (2.5) is available. Begin with W, U i.i.d. $N(0,1)$ random variables and consider the conditional distribution of W given $A = \{\lambda_0+\lambda_1 W > U\}$. However $P(A) = \Phi(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}})$. It follows that

$$f_{W/A}(w;\lambda_0,\lambda_1) = \varphi(w)\int_{-\infty}^{\lambda_0+\lambda_1 w}\varphi(u)du/P(A)$$

which is precisely (2.5).

In the case where $a = \mu_Y$ in (2.2), equivalently when $\lambda_0 = 0$ in (2.5), an alternative stochastic model for (2.5) is possible. (See e.g. Azzalini (1986), Azzalini and Dalla Valle (1996)). If we begin with Y_0 and Y_1 i.i.d. $N(0,1)$ and define

$$W = \delta|Y_0| + \sqrt{1-\delta^2}Y_1 \qquad \ldots (2.6)$$

for $\delta\epsilon(-1,1)$ then the density of W is given by (2.5) with $\lambda_0 = 0$ and $\lambda_1 = \frac{\delta}{\sqrt{1-\delta^2}}$. We have been unable to formulate an extension of the model (2.6) which will lead to the model (2.5) with non-zero values of λ_0.

With different notation the moment generating function of the basic skew-normal density can be found in Arnold *et al.* (1993) and in Chou and Owen (1984). If we denote the moment generating function corresponding to (2.5) by $M(t)$ we may write

$$M(t) \ = \ \int_{-\infty}^{\infty} e^{tw}\varphi(w)\Phi(\lambda_0+\lambda_1 w)\Big/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)dw$$

$$= e^{t^2/2}\int_{-\infty}^{\infty}\varphi(w-t)\Phi(\lambda_0+\lambda_1 w)\Big/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)dw$$

$$[\text{set v=w-t}] \quad = e^{t^2/2}\int_{-\infty}^{\infty}\varphi(v)\Phi(\lambda_0+\lambda_1 t+\lambda_1 v)\Big/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)dv$$

consequently

$$M(t) = e^{t^2/2}\Phi\left(\frac{\lambda_0+\lambda_1 t}{\sqrt{1+\lambda_1^2}}\right)\Big/\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right) . \qquad \ldots(2.7)$$

If W has density (2.5), its moments can be obtained by differentation of the moment generating function (2.7). For example, with $\delta_0 = \lambda_0/\sqrt{1+\lambda_1^2}, \delta_1 = \lambda_1/\sqrt{1+\lambda_1^2}$, and $\psi(\delta_0) = \varphi(\delta_0)/\Phi(\delta_0)$,

$$E(W) = \delta_1\psi(\delta_0) \qquad \ldots(2.8)$$

and

$$var(W) = 1-\delta_0\delta_1^2\psi(\delta_0)-\delta_1^2\psi^2(\delta_0) . \qquad \ldots(2.9)$$

The third standardized cumulant is

$$\gamma_1 = \frac{(\delta_0^2-1)\delta_1^3\psi(\delta_0)-3\delta_0\delta_1^2\psi^2(\delta_0)+2\delta_1^3\psi^3(\delta_0)}{(1-\delta_0\delta_1^2\psi(\delta_0)-\delta_1^2\psi^2(\delta_0))^{3/2}} \qquad \ldots(2.10)$$

For any fixed value of λ_0, when $|\lambda_1| \to \infty$ (i.e. when $|\rho| \to 1$), $\delta_0 \to 0$ and $\delta_1 \to 1$ and $\gamma_1 \to sign(\delta_1)\sqrt{\frac{2}{\pi}}(\frac{4}{\pi}-1)/(1-\frac{2}{\pi})^{3/2}$ which has a maximum of approximately .995 as reported in Azzalini (1986). To obtain higher moments, the recursive relationship among the moments documented in Arnold *et al.* (1993) may be useful.

3. Non-normal Univariate Models

Recall that one formulation of the skew-normal distribution began with W, U i.i.d. $N(0,1)$ random variables and considered the conditional distribution of W given $\lambda_0 + \lambda_1 W > U$. In the case $\lambda_0 = 0$, Arnold and Beaver (1997a) and others suggested models in which U and W had non-normal distributions. Such models can be considered for the case $\lambda_0 \neq 0$ also. Assume that W, U are independent random variables with corresponding density (distribution) functions $\psi_1(\Psi_1)$ and $\psi_2(\Psi_2)$ respectively. Again consider the conditional distribution of W given $\lambda_0+\lambda_1 W > U$. The density of W is given by

$$f(w, \lambda_0, \lambda_1) = \psi_1(w)\Psi_2(\lambda_0+\lambda_1 w)/P(\lambda_0+\lambda_1 W > U) . \qquad \ldots(3.1)$$

The term in the denominator will be generally difficult to evaluate except in special circumstances. Arnold and Beaver (1997a) focussed on the case where $\lambda_0 = 0$ and where both ψ_1 and ψ_2 were symmetric about zero. In this case the denominator simplifies to 1/2 and the density assumes the form

$$f(w; \lambda_1) = 2\psi_1(w)\Psi_2(\lambda_1 w) , \qquad \ldots(3.2)$$

a close parallel to Azzalini's skew-normal density (1.1). When $\lambda_0 \neq 0$, in order that the denominator in (3.1) be computable, it is necessary that $U - \lambda_1 W$ have a "nice" distribution. This will be the case if U, W are independent stable random variables of the same type (for then $U - \lambda_1 W$ will also be stable of the same type). The normal case and the Cauchy case (discussed in Arnold and Beaver (1997b)) are the prime examples. Non-stable tractable cases do exist. One could allow ψ_1 and ψ_2 to be exponential or Laplace densities. (See Balakrishnan and Ambagaspitiye (1994) who considered the skewed Laplace density.) In order to preserve the rationale for calling these skewed distributions, it is reasonable to restrict ψ_1 to be symmetric (though (3.1) makes sense without this restriction), however ψ_2 need not be restricted to be symmetric.

4. Multivariate Hidden Truncation Normal Models

The basic scenario involves a $(k+1)$ dimensional normal random vector $(X_1, \ldots, X_k, Y)$ with a general covariance structure. We then set $\underline{Z} = \underline{X}$ conditional on $Y > a$. Azzalini and Dalla Valle (1996) consider the case $a = \mu_Y$. Here we allow a to be arbitrary. After an affine transformation the distribution of $\underline{Z}$ can be formulated as one obtained as follows (cf. Section 2). Begin with $W_1, \ldots, W_k, U$ i.i.d. $N(0,1)$ random variables and consider the conditional distribution of $\underline{W}$ given that $\lambda_0 + \underline{\lambda}_1' \underline{W} > U$ where $\lambda_0 \epsilon \mathbf{R}$ and $\underline{\lambda}_1 \epsilon \mathbf{R}^k$. The formulation is advantageous since, using it, the form of the resulting distribution is easily found. Note that the conditional density of $(\underline{W}, U)$ given $A = \{\lambda_0 + \underline{\lambda}_1' \underline{W} > U\}$ is

$$f_{\underline{W},U|A}(\underline{w}, u) \propto \prod_{i=1}^{k} \varphi(w_i)\varphi(u) I(\lambda_0 + \underline{\lambda}_1' \underline{w} > u) \ . \qquad \ldots (4.1)$$

If we then integrate with respect to u we get

$$f_{\underline{W}|A}(\underline{w}) = \prod_{i=1}^{k} \varphi(w_i)\Phi(\lambda_0 + \underline{\lambda}_1' \underline{w})/P(A) \ . \qquad \ldots (4.2)$$

However,

$$P(A) = P(U - \underline{\lambda}_1' \underline{W} < \lambda_0) = \Phi(\frac{\lambda_0}{\sqrt{1 + \underline{\lambda}_1' \underline{\lambda}_1}})$$

since $U - \underline{\lambda}_1' \underline{W} \sim N(0, 1 + \underline{\lambda}_1' \underline{\lambda}_1)$. Thus the k-dimensional hidden truncation density analogous to (2.5) is given by

$$f(\underline{w}; \lambda_0, \underline{\lambda}_1) = \prod_{i=1}^{k} \varphi(w_i)\Phi(\lambda_0 + \underline{\lambda}_1' \underline{w}) \Big/ \Phi(\frac{\lambda_0}{\sqrt{1 + \underline{\lambda}_1' \underline{\lambda}_1}}) \ . \qquad \ldots (4.3)$$

The Azzalini and Dalla Valle (1996) k-dimensional density is an affine transformation of (4.3) under the assumption that $\lambda_0 = 0$ (in which case the denominator in

(4.3) takes on the value 1/2). Thus the basic Azzalini and Dalla Valle distribution is of the form

$$f(\underline{w};0,\underline{\lambda}_1) = 2(\prod_{i=1}^{k}\varphi(w_i))\Phi(\underline{\lambda}_1'\underline{w}) \ . \qquad \ldots(4.4)$$

The family of distributions (4.4) clearly has marginals of the same type. As shown below, the larger family (4.3) has marginals and conditionals of the same type.

One way to determine the nature of the marginal distributions of (4.3) is to first identify the moment generating function corresponding to (4.3). We argue as follows:

$$\begin{aligned} M(\underline{t}) &\propto \int\cdots\int_{\mathbf{R}^k} e^{\underline{t}'\underline{u}}(\prod_{i=1}^{k}\varphi(u_i))\Phi(\lambda_0+\underline{\lambda}_1'\underline{u})d\underline{u} \\ &= e^{\underline{t}'\underline{t}/2}\int\int_{\mathbf{R}^k}[\prod_{i=1}^{k}\varphi(u_i-t_i)]\Phi(\lambda_0+\underline{\lambda}_1'\underline{u})d\underline{u} \\ [\text{set } v_i=u_i-t_i] \quad &= e^{\underline{t}'\underline{t}/2}\int\int_{\mathbf{R}^k}[\prod_{i=1}^{k}\varphi(v_i)]\Phi(\lambda_0+\underline{\lambda}_1'\underline{t}+\underline{\lambda}_1'\underline{v})d\underline{v} \\ &= e^{\underline{t}'\underline{t}/2}\Phi(\frac{\lambda_0+\underline{\lambda}_1'\underline{t}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}) \ . \end{aligned}$$

Since $M(\underline{0})=1$, the missing normalizing constant is $1/P(A)$, yielding the moment generating function in the form

$$M(\underline{t}) = e^{\underline{t}'\underline{t}/2}\Phi(\frac{\lambda_0+\underline{\lambda}_1'\underline{t}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})\Big/\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}) \ . \qquad \ldots(4.5)$$

To get the moment generating function of a k_1 dimensional marginal of the distribution (4.3), i.e. of $\dot{\underline{\mathrm{W}}}$ where $\underline{\mathrm{W}} = (\dot{\underline{\mathrm{W}}},\ddot{\underline{\mathrm{W}}})$ has been partitioned into subvectors of dimension k_1 and $k-k_1$, we need only to similarly partition $\underline{\mathrm{t}} = (\dot{\underline{\mathrm{t}}},\ddot{\underline{\mathrm{t}}})$ and set $\ddot{\underline{\mathrm{t}}} = \underline{0}$ in (4.5). It is then obvious that the marginal m.g.f. of $\dot{\underline{\mathrm{W}}}$ is of the same form as (4.5) (though now of dimension k_1).

There is an alternative argument available to justify our claim that the marginals of the density (4.3) are of the same type. As before we will use our dot-double dot notation for partitioning vectors. To get the marginal density of the first k_1 coordinates of (4.3) we want the conditional density of $\dot{\underline{\mathrm{W}}}$ given $\lambda_0+\underline{\lambda}_1'\underline{W}>U$, i.e. given $\lambda_0+\dot{\underline{\lambda}}_1'\dot{\underline{W}} > U-\ddot{\underline{\lambda}}_1'\ddot{\underline{W}}$. But the random variable $U-\ddot{\underline{\lambda}}_1'\ddot{\underline{W}} \sim N(0,1+\ddot{\underline{\lambda}}_1'\ddot{\underline{\lambda}}_1)$ so that our conditioning event can be written as

$$\frac{\lambda_0}{\sqrt{1+\ddot{\underline{\lambda}}_1'\ddot{\underline{\lambda}}_1}} + \frac{\dot{\underline{\lambda}}_1'\dot{\underline{W}}}{\sqrt{1+\ddot{\underline{\lambda}}_1'\ddot{\underline{\lambda}}_1}} > \tilde{V}$$

where $\tilde{V}$ is $N(0,1)$ and is independent of $\dot{\underline{W}}$. Then the k_1-dimensional marginal is of the form

$$f(\underline{\dot{w}}) = \frac{[\prod_{i=1}^{k_1} \varphi(w_i)]\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\ddot{\lambda}}_1'\underline{\ddot{\lambda}}_1}} + \frac{\underline{\dot{\lambda}}_1'\underline{\dot{w}}}{\sqrt{1+\underline{\ddot{\lambda}}_1'\underline{\ddot{\lambda}}_1}})}{\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})} \qquad \ldots(4.6)$$

By analogous arguments, the marginal density of the last $k - k_1$ coordinates of $\underline{W}$ in (4.3) is

$$f(\underline{\ddot{w}}) = \frac{[\prod_{i=k_1+1}^{k} \varphi(w_i)]\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}} + \frac{\underline{\ddot{\lambda}}_1'\underline{\ddot{w}}}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}})}{\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})} \qquad \ldots(4.7)$$

and consequently the conditional density of $\underline{\dot{W}}$ given $\underline{\ddot{W}} = \underline{\ddot{w}}$ is (taking the ratio of (4.3) and (4.7))

$$f(\underline{\dot{w}}|\underline{\ddot{w}}) = \frac{[\prod_{i=1}^{k_1} \varphi(w_i)]\Phi(\lambda_0 + \underline{\ddot{\lambda}}_1'\underline{\ddot{w}} + \underline{\dot{\lambda}}_1'\underline{\dot{w}})}{\Phi(\frac{\lambda_0+\underline{\ddot{\lambda}}_1'\underline{\ddot{w}}}{\sqrt{1+\underline{\dot{\lambda}}_1'\underline{\dot{\lambda}}_1}})} . \qquad \ldots(4.8)$$

Thus all conditionals as well as all marginals of the density (4.3) are of the same type.

When the $(k+1)$ dimensional normal random vector $(X_1, \ldots, X_k, Y)$ has a mean $\underline{\mu}$ and covariance matrix $\Sigma > 0$, both the marginal and conditional distributions belong to the same family of multivariate skewed normal distributions. Beginning with (4.3), consider the transformation $\underline{W} = \Sigma^{-\frac{1}{2}}(\underline{X} - \underline{\mu})$ to find that

$$f_{\underline{X}}(\underline{x}) \propto \exp\left(-\frac{1}{2}(\underline{x} - \underline{\mu})'\Sigma^{-1}(\underline{x} - \underline{\mu})\right) \times \Phi\left(\frac{\delta_0}{\sqrt{1 - \underline{\delta}_1'\underline{\delta}_1}} + \frac{\underline{\delta}_1'}{\sqrt{1 - \underline{\delta}_1'\underline{\delta}_1}}\Sigma^{-\frac{1}{2}}(\underline{x} - \underline{\mu})\right) \qquad \ldots(4.9)$$

with $\delta_0 = \lambda_0/\sqrt{1 + \underline{\lambda}_1'\underline{\lambda}_1}$ and $\underline{\delta}_1 = \underline{\lambda}_1/\sqrt{1 + \underline{\lambda}_1'\lambda_1}$. Now consider partitioning $\underline{x}' = (\underline{\dot{x}}', \underline{\ddot{x}}')$ with a corresponding partition of Σ and $\underline{\lambda}_1$. Using the moment-generating function, the marginals are of the form

$$f_{\underline{\ddot{X}}}(\underline{\ddot{x}}) \propto \exp\left(-\frac{1}{2}(\underline{\ddot{x}} - \underline{\ddot{\mu}})'\Sigma_{22}^{-1}(\underline{\ddot{x}} - \underline{\ddot{\mu}})\right) \times \Phi\left(\frac{\delta_0}{\sqrt{1 - \underline{\ddot{\delta}}_1'\underline{\ddot{\delta}}_1}} + \frac{\underline{\ddot{\delta}}_1'}{\sqrt{1 - \underline{\ddot{\delta}}_1'\underline{\ddot{\delta}}_1}}\Sigma_{22}^{-\frac{1}{2}}(\underline{\ddot{x}} - \underline{\ddot{\mu}})\right). \qquad \ldots(4.10)$$

The conditional density of $\underline{\dot{X}}|\underline{\ddot{X}}$ is

$$f_{\underline{\dot{X}}|\underline{\ddot{X}}}(\underline{\dot{x}}|\underline{\ddot{x}}) \propto \exp\left(-\frac{1}{2}(\underline{\dot{x}} - \underline{\dot{\mu}}(\underline{\ddot{x}})')(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(\underline{\dot{x}} - \underline{\dot{\mu}}(\underline{\ddot{x}}))\right)$$

$$\times\Phi\left(\frac{\delta_0}{\sqrt{1 - \underline{\delta}_1'\underline{\delta}_1}} + \frac{\underline{\delta}_1'}{\sqrt{1 - \underline{\delta}_1'\underline{\delta}_1}}\Sigma^{-\frac{1}{2}}(\underline{x} - \underline{\mu})\right) \qquad \ldots(4.11)$$

for $\underline{\dot{\mu}}(\underline{\ddot{x}}) = \underline{\dot{\mu}}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underline{\ddot{x}} - \underline{\ddot{\mu}})$. In the second term in (4.11), partition the $k \times k$ matrix $\Sigma^{-\frac{1}{2}}$ as $[(\Sigma^{-\frac{1}{2}})_{(1)}, (\Sigma^{-\frac{1}{2}})_{(2)}]$ where $(\Sigma^{-\frac{1}{2}})_{(1)}$ and $(\Sigma^{-\frac{1}{2}})_{(2)}$ are of dimensions $k \times k_1$ and $k \times k_2$, respectively. This term can be written as

$$\Phi\left(\frac{\delta_0}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}} + \frac{\underline{\delta}_1'}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}}(\Sigma^{-\frac{1}{2}})_{(2)}(\underline{\ddot{x}} - \underline{\ddot{\mu}}) + \frac{\underline{\delta}_1'}{\sqrt{1-\underline{\delta}_1'\underline{\delta}}}(\Sigma^{-\frac{1}{2}})_{(1)}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{\frac{1}{2}} \times (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-\frac{1}{2}}(\underline{\dot{x}} - \underline{\dot{\mu}})\right), \quad \ldots(4.12)$$

which is of the same form with $\tau_0 = \delta_0(\underline{\ddot{x}})$ and $\underline{\tau}_1 = \underline{\delta}_1(\underline{\ddot{x}})$ where

$$\frac{\tau_0}{\sqrt{1-\underline{\tau}_1'\underline{\tau}_1}} = \frac{\delta_0}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}} + \frac{\underline{\delta}_1'}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1'}}(\Sigma^{-\frac{1}{2}})_{(2)}(\underline{\ddot{x}} - \underline{\ddot{\mu}}) \quad \ldots(4.13)$$

and

$$\frac{\underline{\tau}_1'}{\sqrt{1-\underline{\tau}_1'\underline{\tau}}} = \frac{\underline{\delta}_1'}{\sqrt{1-\underline{\delta}_1'\underline{\delta}_1}}(\Sigma^{-\frac{1}{2}})_{(1)}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{\frac{1}{2}}\ . \quad \ldots(4.14)$$

Therefore, the family of multivariate skewed normal distributions (4.9) is closed under marginalization and conditioning.

Moments of the distribution (4.3) and of its conditional distributions are obtained by differentiating the moment generating function or by differentiating the logarithm of the moment generating function. From (4.5), we have

$$\log M(t) = \sum_{i=1}^{k} t_i^2 + \log\Phi\left(\frac{\lambda_0 + \underline{\lambda}_1'\underline{t}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right) - c\ .$$

Differentiating with respect to t_i and setting $\underline{t} = \underline{0}$ yields

$$E(W_i) = \frac{\lambda_{1i}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}} \frac{\varphi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}{\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}\ . \quad \ldots(4.15)$$

Further differentiation yields the variances and covariances.

$$cov(W_i, W_j) = \delta_{ij} - \frac{\lambda_{1i}\lambda_{1j}}{(1+\underline{\lambda}_1'\underline{\lambda}_1)}\left[\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}} g(\lambda_0, \underline{\lambda}_1) + g^2(\lambda_0, \underline{\lambda}_1)\right] \quad \ldots(4.16)$$

where δ_{ij} is the Kronecker delta symbol and where

$$g(\lambda_0, \underline{\lambda}_1) = \varphi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right) \Big/ \Phi\left(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}\right)\ . \quad \ldots(4.17)$$

Naturally, considerable simplification occurs when $\lambda_0 = 0$, in which case the expression in (4.17) simplifies to $\sqrt{\frac{2}{\pi}}$.

Azzalini and Dalla Valle (1996) remark that in the case $\lambda_0 = 0, \sum_{i=1}^k W_i^2$ has a χ_k^2 distribution. This can be confirmed by considering the moment generating function of $\sum_{i=1}^k W_i^2$ in the general case where $\underline{W}$ has density given by (4.3). We find

$$E(e^{t\sum_{i=1}^k W_i^2}) = \int\cdots\int_{\mathbf{R}^k} e^{t\underline{w}'\underline{w}}[\prod_{i=1}^k \varphi(w_i)]\frac{\Phi(\frac{\lambda_0+\underline{\lambda}_1'\underline{w}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}{\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}d\underline{w}$$

$$[v_i = (1-2t)^{1/2}w_i] \quad = (1-2t)^{-k/2}\int\cdots\int_{\mathbf{R}^k}[\prod_{i=1}^k \varphi(v_i)]\frac{\Phi(\frac{\lambda_0+(1-2t)^{-1/2}\underline{\lambda}_1'\underline{v}}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}{\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}})}d\underline{v}$$

$$= (1-2t)^{-k/2}\Phi(\frac{\lambda_0}{\sqrt{1+\frac{\underline{\lambda}_1'\underline{\lambda}_1}{1-2t}}})\Big/\Phi(\frac{\lambda_0}{\sqrt{1+\underline{\lambda}_1'\underline{\lambda}_1}}) \ .$$

This reduces to the χ_k^2 moment generating function ($(1-2t)^{-k/2}$ only when $\lambda_0 = 0$. In general, then, $\sum_{i=1}^k W_i^2$ will not have a χ_k^2 distribution.

5. **Non-normal Multivariate Models**

If we dispense with the assumption that the basic random variables $W_1, W_2, \ldots, W_k, U$, discussed in the last section, are normal then we can expect to encounter new kinds of skewed multivariate distributions. In full generality we consider the case where $W_1, W_2, \ldots, W_k$ and U are independent random variables with densities (distributions) given by $\psi_1(\Psi_1), \psi_2(\Psi_2), \ldots, \psi_k(\Psi_k), \psi_0(\Psi_0)$ respectively. We then consider the conditional distribution of $\underline{W}$ given that $\lambda_0+\underline{\lambda}_1'\underline{W} > U$, as before, where $\lambda_0\epsilon\mathbf{R}$ and $\underline{\lambda}\epsilon\mathbf{R}^k$. The conditional density of $(\underline{W}, U)$ given $A = \{\lambda_0 + \lambda_1\underline{W} > U\}$ is of the form

$$f_{\underline{W},U|A}(\underline{w}, u) \propto [\prod_{i=1}^k \psi_i(w_i)]\psi_0(u)I(\lambda_0 + \underline{\lambda}_1'\underline{w} > u) \ . \qquad \ldots(5.1)$$

Integrating with respect to u yields

$$f_{\underline{W}|A}(\underline{w}) = [\prod_{i=1}^k \psi_i(w_i)]\Psi_0(\lambda_0 + \underline{\lambda}_1'\underline{w})/P(A) \ . \qquad \ldots(5.2)$$

As in the univariate case, the evaluation of $P(A)$ can be troublesome. An exception occurs if $\lambda_0 = 0$ and if each of the densities $\psi_1, \ldots, \psi_k, \psi_0$ is symmetric. For in that case $U - \underline{\lambda}'\underline{W}$ is a symmetric random variable and $P(A) = 1/2$. This leads to the multivariate version of (3.2) namely

$$f(\underline{w}) = 2[\prod_{i=1}^k \psi_i(w_i)]\Psi_0(\underline{\lambda}_1'\underline{w}) \qquad \ldots(5.3)$$

where $\psi_1, \ldots, \psi_k$ are symmetric densities and Ψ_0 is a symmetric distribution function. The Azzalini and Dalla Valle (1996) k-variate skew-normal distribution (4.4) is of the form (5.3) with $\psi_1 = \psi_2 = \ldots = \psi_k = \psi_0 = \varphi$, the standard normal density. By choosing the ψ_i's in (5.3) to be identical but not normal, leads to natural analogs of the density (4.4) which respectively can be called skew-Cauchy, skew-Laplace, skew-logistic, etc.

If $\lambda_0 \neq 0$ in (5.2), then computation of $P(A)$ can rarely be accomplished. An important exception occurs when each of the densities of $W_1, W_2, \ldots, W_k$ and U are stable (and of the same type). For then, $U - \underline{\lambda}'\underline{W}$ will also have a stable distribution. The normal case is a natural example. So also is the Cauchy. For the basic k-variate skew-Cauchy distribution we begin with $W_1, W_2, \ldots, W_k, U$ i.i.d. standard Cauchy $(0,1)$ random variables and consider the conditional distribution of W given $\lambda_0 + \underline{\lambda}_1\underline{W} > U$. Since $U - \underline{\lambda}_1\underline{W} \sim C(0, 1 + \sum_{i=1}^k |\lambda_{1i}|)$, we obtain, using (5.2), the density

$$f(\underline{w}) = [\prod_{i=1}^{k} \psi^*(w_i)]\Psi^*(\lambda_0 + \underline{\lambda}_1'\underline{w})/\Psi^*(\frac{\lambda_0}{1 + \sum_{i=1}^k |\lambda_{1i}|}) \qquad \ldots (5.4)$$

where

$$\psi^*(w) = \frac{1}{\pi(1+w^2)}, \quad w \epsilon \mathbf{R}$$

and

$$\Psi^*(w) = \frac{1}{2} + \frac{1}{\pi} Tan^{-1}(w), \quad w \epsilon \mathbf{R}\ ,$$

the standard Cauchy density and distribution respectively. For more detailed discussion of the skew-Cauchy density (5.4) see Arnold and Beaver (1997b). Note that the skew-Cauchy density (5.4) has all marginals and conditionals in the same family (a parallel to the situation encountered with the skew-normal density (4.3)).

6. **Multiple Constraint Models**

In addition to the example of weights of police officers selected on standing and sitting heights, we can also consider situations such as admission to a Graduate School which is contingent upon an applicant's GPA and GRE scores (V = verbal, Q = quantitative, A = analytical). With W_1 = GPA, W_2 = V, W_3 = Q and W_4 = A, unconditional admission may require that $W_1 > 3.2$, $W_2 + W_3 > 1100$ and $W_2 + W_3 + W_4 > 1800$. Hence we are led to consider multiple constraint extensions of the hidden truncation models introduced in Sections 2-5. The normal multiple constraint model will be most tractable but there is no reason not to begin with a more general model involving non-normal, not necessarily identical basic distributions (analogous to the cases in Section 5).

Thus we begin with $k + \ell$ independent random variables $W_1, W_2, \ldots, W_k$, $U_1, U_2, \ldots, U_\ell$. Then we consider the conditional distribution of $\underline{W}$ given that for

$j = 1, 2, \ldots, \ell$, $\lambda_0^{(j)} + \underline{\lambda}_1^{(j)} \underline{W} > U_j$. We will denote the density (distribution) functions of the W_i's by $\psi_i(\Psi_i)$ respectively and the density (distribution) functions of the U_j's by $\tilde{\psi}_j(\tilde{\Psi}_j)$ respectively. Note that we allow the possibility that the $k + \ell$ distributions will be distinct; though we may expect to encounter some simplification in cases where they are more homogeneous in type (e.g. if all are normal). For notational convenience we define events $A_j, j = 1, 2, \ldots, \ell$ by

$$A_j = \{\lambda_0^{(j)} + \underline{\lambda}_1^{(j)} \underline{W} > U_j\} \ . \qquad \ldots (6.1)$$

Now we first consider the conditional density of $(\underline{W}, \underline{U})$ given $A^* = \cap_{j=1}^{\ell} A_j$, i.e.

$$f_{\underline{W},\underline{U}|A^*}(\underline{w}, \underline{u}) = \frac{[\prod_{i=1}^{k} \psi_i(w_i)][\prod_{j=1}^{\ell} \tilde{\psi}_j(u_j)](\prod_{j=1}^{\ell} I(\lambda_0^{(j)} + \underline{\lambda}_1^{(j)'} \underline{w} > u_j))}{P(\cap_{j=1}^{\ell} A_j)} \ . \qquad \ldots (6.2)$$

Next, integrate out $\underline{u}$ to get the desired conditional density

$$f_{\underline{W}|A^*}(\underline{w}) = \frac{[\prod_{i=1}^{k} \psi_i(w_i)] \prod_{j=1}^{\ell} \Psi_j(\lambda_0^{(j)} + \underline{\lambda}_1^{(j)'} \underline{w})}{P(\cap_{j=1}^{\ell} A_j)} \ . \qquad \ldots (6.3)$$

As in the single constraint case, it is the denominator in (6.3) which generally will be difficult to deal with.

One case in which the denominator is computable is that in which all the ψ_i's and $\tilde{\psi}_j$'s are standard normal (denoted as usual by φ) <u>and</u> where the skewness vectors $\underline{\lambda}_1^{(j)}, j = 1, 2, \ldots, \ell$ are mutually orthogonal. In such a setting it is clear that the A_j's (defined in (6.1)) are independent events. Thus if the $\underline{\lambda}_1^{(j)}$'s are orthogonal and the basic densities are normal we have

$$f_{\underline{W}|A^*}(\underline{w}) = \frac{[\prod_{i=1}^{k} \varphi(w_i)] \prod_{j=1}^{\ell} \Phi(\lambda_0^{(j)} + \underline{\lambda}_1^{(j)'} \underline{w})}{\prod_{j=1}^{\ell} \Phi(\frac{\lambda_0^{(j)}}{\sqrt{1 + \underline{\lambda}_1^{(j)'} \underline{\lambda}_1^{(j)}}})} \ . \qquad \ldots (6.4)$$

Even more simplification will be encountered if $\lambda_0^{(j)} = 0, j = 1, 2, \ldots, \ell$. In such a case

$$f_{\underline{W}|A^*}(\underline{w}) = 2^k [\prod_{i=1}^{k} \varphi(w_i)] \prod_{j=1}^{\ell} \Phi(\underline{\lambda}_1^{(j)'} \underline{w}) \ , \qquad \ldots (6.5)$$

recognizable as a natural extension of the Azzalini and Dalla Valle k-variate skew normal density (4.4).

7. The Australian Athletes Revisited

In this section, we shall use the data set reported by Cook and Weisberg (1994), concerning thirteen variables measured on 202 athletes at the Australian Institute of Sport, courtesy of Richard Telford and Ross Cunningham. We shall consider the person's height and weight, denoted by (H, W) as the pair of variables to be analyzed using a model based upon (1.3) in which (1.3) is given in terms of the transformed variables $\underline{x}$. Here $\underline{u} = \sum^{-1/2}(\underline{x} - \underline{\mu})$ with

$$\Sigma^{-1/2} = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{bmatrix} \qquad \ldots (7.1)$$

The parameterization in (7.1), which was used for computational simplicity, reflects the natural parameterization in an exponential family of distributions. The method of maximum likelihood was implemented using the Matlab minimization routine. Values obtained using a genetic algorithm and simulated annealing were used as starting values for the Matlab routine. The maximum likelihood estimates together with the value of the loglikelihood for the 5- and 7- parameter distributions based upon the bivariate normal are given in Table 1.

Table 1. The loglikelihood and parameter estimates for 5, 7 and 8 parameter models

	Models		
Parameters Estimated	5-parameter	7-parameter	8-parameter
λ_0	-	-	-4.2050
λ_1	-	-.8617	-.9816
λ_2	-	3.1181	3.8888
μ_1	180.10	178.34	176.92
μ_2	75.01	63.39	46.09
θ_1	.1541	.1375	.1281
θ_2	-.0586	-.0381	-.0264
θ_3	.0993	.0678	.0523
Loglikelihood	-1468.85	-1453.75	-1451.75

In testing whether the 7-parameter model produces a better fit than the 5-parameter model, the likelihood ratio statistic, $\Lambda = f(\underline{x} : \hat{\lambda}_0, \hat{\lambda}_1, \hat{\underline{\mu}}, \hat{\underline{\theta}})/f(\underline{x} : \hat{\underline{\mu}}, \hat{\underline{\theta}})$ produced a value of $-2ln(\Lambda) = -2[(-1468.85) - (1453.75)] = 30.20$, which when compared to critical values of a χ^2-distribution with 2 degrees of freedom has a p-value $\leq .0001$. Hence, the 7-parameter model produces a significantly better fit.

The results of an analysis of the data based upon an 8-parameter model as in (1.3) in which an additional truncation parameter, λ_0, is included are also given in Table 1.

Assessing the fit of the 8-parameter model compared to the 7-parameter model produced a value of minus twice the log of the likelihood ratio statistic equal to 4.00. When compared to the critical values of a χ^2 distribution with one degree of freedom the likelihood-ratio statistic produced $.01 \leq$ p-value $\leq .05$, indicating that the additional truncation parameter is significant. In both cases, it appears that the data have come from a distribution in which hidden truncation was used in data collection.

Acknowledgements. The authors would like to thank Mark E. Lehr for his role in analyzing the data. The genetic algorithm and the simulated annealing routines

used in maximizing the likelihoods were written by him and are available by e-mailing him at mlehr@pacbell.net.

References

Arnold, B.C., Beaver, R.J., Groeneveld, R.A. and Meeker, W.Q. (1993). The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-478.

Arnold, B.C. and Beaver, R.J. (1997a). Some skewed multivariate models. *Technical Report* # 249, Department of Statistics, University of California, Riverside.

− − −− (1997b). The multivariate skew-Cauchy distribution. *Technical Report* #250, Department of Statistics, University of California, Riverside.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

− − −− (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **56**, 199-208.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Balakrishnan, N. and Ambagaspitiya, R.S. (1994). On skew-Laplace distributions. *Technical Report*, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Commun. Statist. - Theory Meth.*, 19(1), 197-203.

Chou, Y-M. and Owen, D.B. (1984). An Approximation to the percentiles of a variable of the bivariate normal distribution when the other variable is truncated, with applications. *Comm. Statist. - Theory Meth.* **13**, 2535-2547.

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics.* New York: Wiley

Barry C. Arnold and Robert J. Beaver
Department of Statistics
University of California
Riverside, CA 92521
e-mail : barry.arnold@ucr.edu, beaver@ucrac1.ucr.edu